Assignment 11

Ariful Islam

July 2025

Abstract

This report presents a comparative study of face verification models trained using Binary Cross-Entropy (BCE), Contrastive, and Triplet loss functions. Additionally, the impact of using BCE and Mean Squared Error (MSE) as reconstruction losses in a Variational Autoencoder (VAE) is investigated. All models were trained and evaluated on the Labeled Faces in the Wild (LFW) dataset, and their performances are analyzed in terms of accuracy, error rates, and visual outputs.

1 Problem Statement

1. comparing three face verifiers which are trained using Binary Cross-Entropy loss, contrastive loss and triplet loss, receptively.
2. comparing effect of Binary Cross-Entropy (BCE) loss and Mean Squared Error (MSE) as a reconstruction loss during training of a Variational Autoencoder (VAE).

2 Introduction

Loss functions play a critical role in training deep learning models, especially in tasks such as face verification and image reconstruction. This report aims to compare:

- The performance of three face verification models using different loss functions.
- The effect of two reconstruction losses (BCE vs MSE) in training a Variational Autoencoder (VAE).

Understanding how each loss influences learning helps in selecting the appropriate method for a given task.

3 Face Verifier Comparison

3.1 Dataset and Preprocessing

We used the Labeled Faces in the Wild (LFW) dataset and selected 2000 face images. Each image was resized to $128 \times 128 \times 3$, normalized to [0, 1], and paired into:

- **Positive pairs**: same person
- **Negative pairs**: different people

Pairs were split into training, validation, and test sets in 72%, 18%, and 10% ratios, respectively.

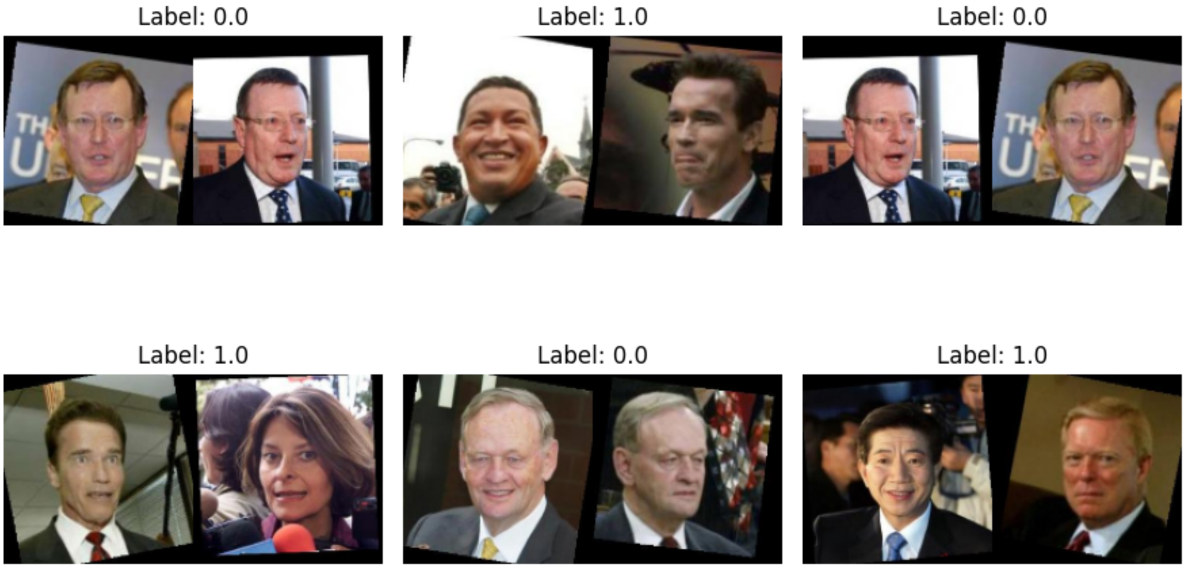

Figure 1: Demo image from dataset with ground truth

3.2 Model Architecture and Loss Functions

All face verification models share a common CNN architecture that outputs embeddings. The final similarity is computed differently based on the loss:

- **Binary Cross-Entropy (BCE)**: Predicts binary labels (same/different).
- **Contrastive Loss**: Optimizes distance between embeddings based on class similarity.
- **Triplet Loss**: Uses (anchor, positive, negative) images to learn relative distances.

3.3 BCE

The face verification model trained with Binary Cross-Entropy (BCE) loss showed steady improvements in training accuracy, reaching up to **97.18%** by epoch 20. However, the validation accuracy plateaued around **46.88%**, indicating signs of overfitting. The final validation loss was approximately **0.9938**, despite a significant reduction in training loss to **0.2580**. This suggests that while the model was able to fit the training data well, its generalization to unseen pairs remained limited. The architecture consisted of a shared embedding network followed by a distance-based comparison and a binary output layer, totaling approximately **190K trainable parameters**.

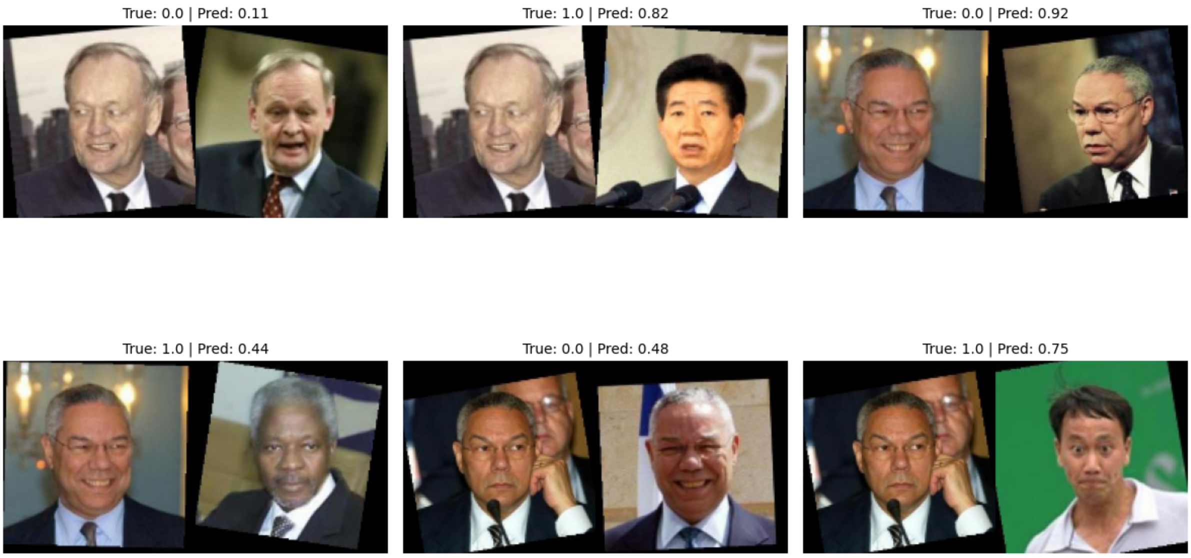

Figure 2: The figure shows the protection of face varifire in BCE

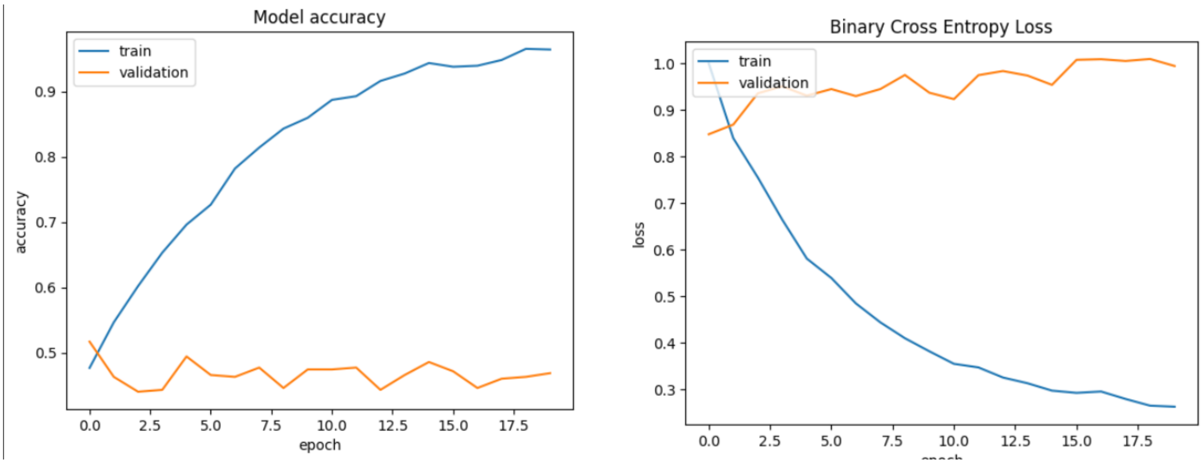

Figure 3: Model accuracy and BCE loss function graph

3.4 Contrastive Loss

The model trained using Contrastive loss demonstrated gradual improvement in both training and validation performance. Training accuracy increased to approximately **68.91%** by epoch 20, while validation accuracy peaked at around **67.61%**, with a corresponding minimum validation loss of **0.2070**. Unlike BCE loss.

Table 1: Performance Metrics for Contrastive Loss Model

Metric	**Value**
Training Accuracy	68.91%
Validation Accuracy	67.61%
Validation Loss	0.2070
Epochs Trained	20

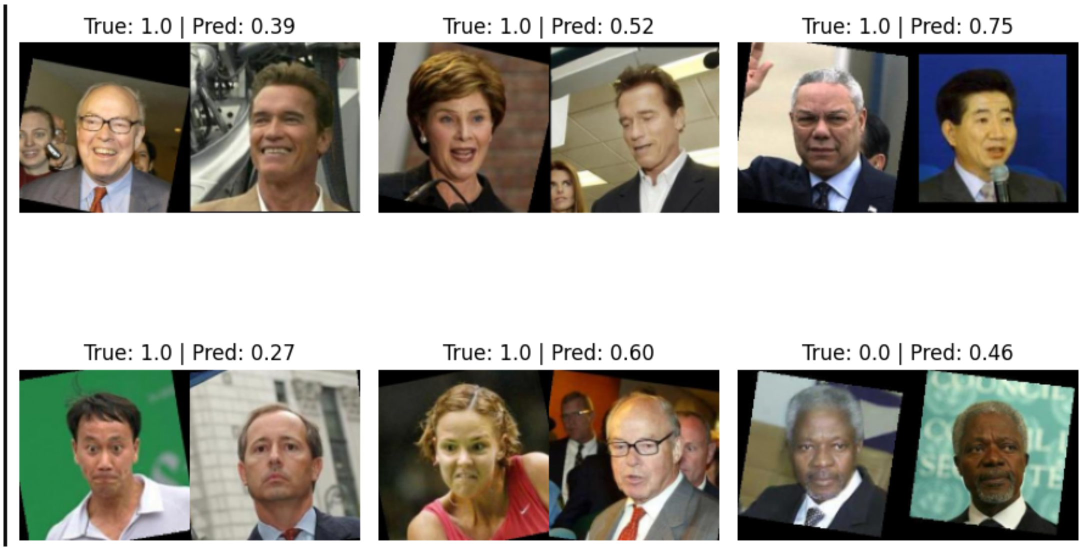

Figure 4: Prediction Batch of Contrastive Loss

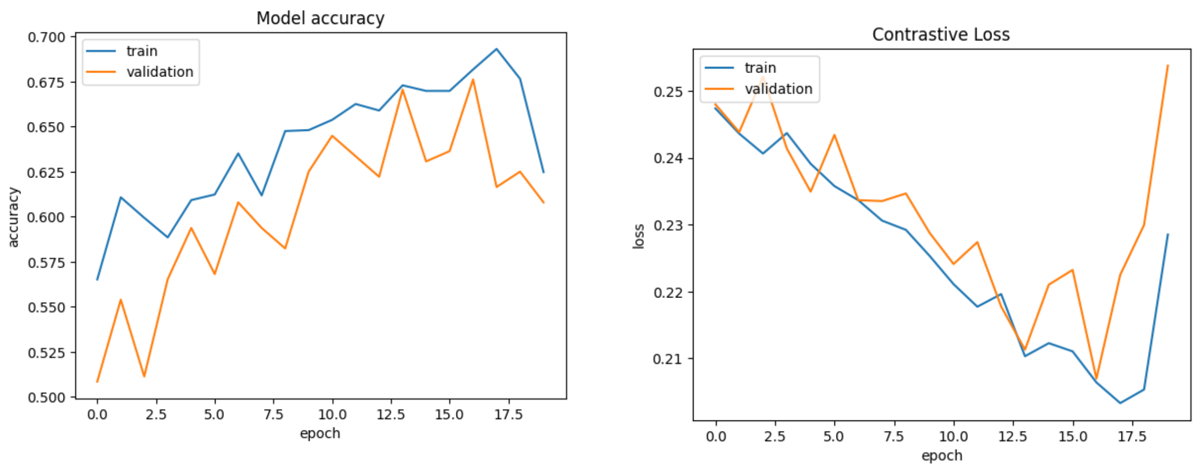

Figure 5: The model accuracy and loss function after training with contrastive loss

3.5 Triplet Loss

The Triplet Loss model shows slow improvement in training loss, decreasing from 0.35 to 0.02 over 20 epochs. Although the validation loss fluctuated between 0.25 and 0.35, it remained relatively stable, indicating that the model learned meaningful embedding relationships while showing some signs of overfitting toward the later epochs.

Table 2: Summary Metrics for Triplet Loss Model after 20 Epochs

Metric	Value
Final Training Loss	0.0217
Final Validation Loss	0.3372
Epochs Trained	20

Figure 6: Prediction Batch of Triplet Loss

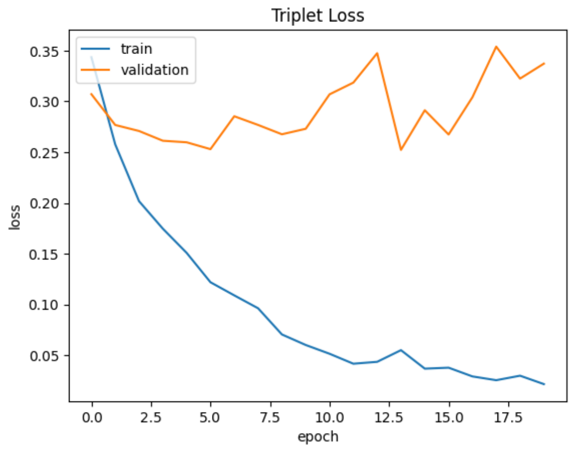

Figure 7: Training and validation loss

3.6 Results Analysis

The comparison says that Binary Cross-Entropy (BCE) loss leads to high training accuracy (97.18%) but poor generalization with low validation accuracy (46.88%), indicating overfitting. Contrastive Loss provides a balanced result with good generalization and validation accuracy (67.61%), making it more reliable. Triplet Loss achieves the lowest training loss (0.0217) and stable validation loss, though accuracy metrics were not explicitly recorded. Overall, Triplet and Contrastive losses outperform BCE for face verification tasks due to better embedding learning and generalization.

Table 3: Comparison of Face Verifier Models Trained with Different Loss Functions

Metric	BCE Loss	Contrastive Loss	Triplet Loss
Training Accuracy	97.18%	68.91%	—
Validation Accuracy	46.88%	67.61%	—
Final Training Loss	0.2580	—	0.0217
Final Validation Loss	0.9938	0.2070	0.3372
Epochs Trained	20	20	20
Overfitting Observed	Yes	No	Slight

4 Task 2

A Variational Autoencoder (VAE) learns to encode and reconstruct images by minimizing the combination of a reconstruction loss and a KL-divergence term. We experimented with two reconstruction losses:

- **Binary Cross-Entropy (BCE)**: Treats pixel values as probabilities.
- **Mean Squared Error (MSE)**: Minimizes pixel-wise squared differences.

4.1 Dataset

The experiment used the MNIST dataset, a widely recognized benchmark containing 70,000 grayscale images of handwritten digits (0–9), each sized 28×28. All images were converted into tensors and normalized to [0,1] using the ToTensor() transform. The dataset was split into 60,000 training and 10,000 test images and loaded in batches using PyTorch's DataLoader.

4.2 Training Configuration

The Variational Autoencoder (VAE) was trained on the MNIST dataset for 10 epochs using both Binary Cross-Entropy (BCE) and Mean Squared Error (MSE) as reconstruction losses.

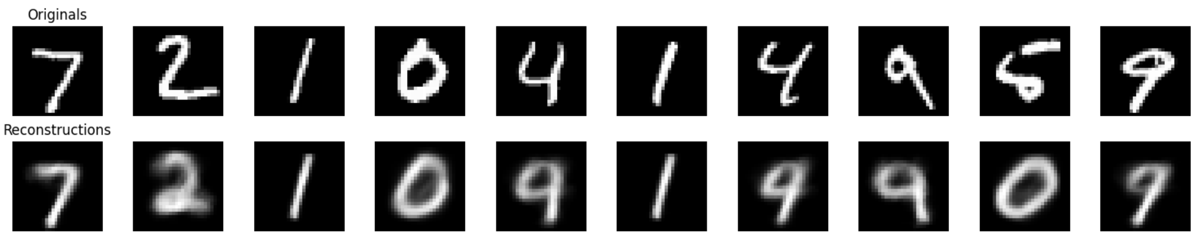

Figure 8: Reconstruction using BCE loss

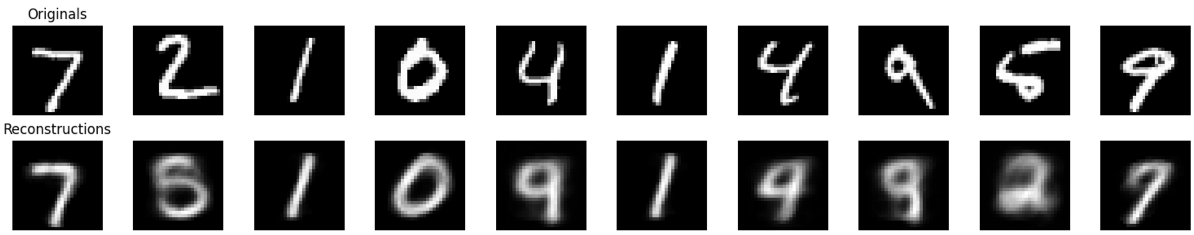

Figure 9: Reconstruction using MSE loss

4.3 Results

Both BCE and MSE produced plausible reconstructions, but with key differences:

- **BCE Loss**: Sharper reconstructions but more prone to overfitting. With BCE, the total loss decreased from 192.09 to 156.09.
- **MSE Loss**: Smoother reconstructions, better generalization, slightly blurrier output. MSE-based training reduced the loss from 48.84 to 39.34.

MSE converged faster and more smoothly, indicating better training stability on this pixel-intensity-based dataset

5 Conclusion

In this report, I try to describe how loss function selection is crucial for both face verification and generative modeling. Triplet loss is ideal for high-accuracy verification systems, while MSE is recommended for robust VAE training.

6 Source Code Link

GitHub Repository